Optimal Kernel Estimation of Spot Volatility of Stochastic Differential Equations

José E. Figueroa-López*
Department of Mathematics, Washington University in St. Louis
and
Cheng Li
Department of Statistics, Purdue University

February 16, 2018

Abstract

A feasible method of bandwidth and kernel selection for spot volatility kernel estimators is proposed, under some mild conditions on the volatility process, which not only cover classical Brownian motion driven dynamics but also some processes driven by long-memory fractional Brownian motions. We characterize the leading order terms of the mean squared error, which in turn enables us to determine an explicit formula for the leading term of the optimal bandwidth. Central limit theorems for the estimation error are also obtained. A feasible plug-in type bandwidth selection procedure is then proposed, for which, as a sub-problem, a new estimator of the volatility of volatility is developed. The optimal selection of the kernel function is also investigated. For Brownian Motion type volatilities, the optimal kernel turns out to be an exponential function, while, for fractional Brownian motion type volatilities, numerical results to compute the optimal kernel are devised. Simulation studies further confirm the good performance of the proposed methods.

Keywords: Spot volatility estimation; Kernel estimation; Bandwidth selection; Kernel function selection; vol vol estimation.

*The first author's research is partially supported by the NSF grants DMS-1149692 and DMS-1613016.

1 Introduction

The estimation of the diffusive coefficient σ_t of the dynamical stochastic system $dX_t = \mu_t dt + \sigma_t dW_t$, driven by a Brownian motion W, has received some renewed attention in the last few years. This research has partly been pushed by the advent of high-frequency data (HFD) in several fields but more predominantly in finance. In the latter context, σ_t is called the spot volatility of the price process $S_t = \exp(X_t)$ of a risky asset and, in addition of being a local measure of the asset's riskiness at the time t, it is also needed for many problems of finance such as option pricing and portfolio selection.

In this work, we revisit the problem of spot volatility estimation by kernel methods. Kernel estimation has a long history and extensive treatments of the method can be found in many textbooks. The selection of the bandwidth and the kernel function are of great importance for the performance of the kernel estimator in a finite sample setting. The problem has been extensively studied for density estimation and kernel regression (cf. Park & Marron (1990), Jones et al. (1996), Epanechnikov (1969)). However, in the context of spot volatility estimation, the literature related to this problem is much scarcer. In this work, we put forward a unified framework to the problem that allows us to deal not only with well studied Brownian driven volatilities but also those driven by other Gaussian processes such as fractional Brownian motions.

Literature review. Foster & Nelson (1994) studied a rolling window estimator, which can be seeing as a kernel estimator with a compactly supported kernel function. They established the point-wise asymptotic normality of the estimator, and drew some conclusions about the optimal window length (i.e., bandwidth) and the optimal weight functions (kernel functions). However, in spite of the non-parametric model setting, the volatility was constrained to have a Brownian-like degree of smoothness (see Assumption A (vii) and

(viii) therein) and the selection of bandwidth and kernel function was not systematically studied, since it was assumed the strict relationship[1] $h_n \asymp n^{-1/2}$ between the window's length h_n and the sample size n (see Assumption D therein). Under such a relationship, they obtained the optimal kernel weights and separately determine the optimal constant c appearing in the formula $h_n = cn^{-1/2}$, but only for the flat-weights or uniform kernel case (see Theorem 4 therein). Fan & Wang (2008) also showed a point-wise asymptotic normality for a general kernel estimator under a specific constraint on the rate of convergence of the bandwidth (Condition A4 therein), which allowed them to neglect the error coming from approximating the spot volatility by a kernel weighted volatility (we refer the reader to Section 6 for details), but the achieved convergence rates are suboptimal. For a continuous Itô semimartingale with volatility driven by a Brownian motion and jumps, Alvarez et al. (2012) considered the estimation of σ_t^p by taking forward finite differences of the realized power variation process of order p, which is equivalent to a forward-looking kernel estimator with uniform kernel. CLTs were also developed therein, which allowed them to argue that the best possible rate of convergence of the estimation error is $n^{-1/4}$ and that this is attained when $n^{1/2}h_n \to c \in (0, \infty)$, as $n \to \infty$. More general results along the same vein have also been developed in the monograph of Jacod & Protter (2012) (see Chapter 13 therein). More recently, Mancini et al. (2004) has developed asymptotic normality for a more general class of spot volatility estimators, which includes kernel estimators.

Besides Foster & Nelson (1994), the only work we know that studied the problem of bandwidth selection of spot volatility kernel estimators is that of Kristensen (2010), who also obtained asymptotic normality of the estimators. However, this work imposes a strong path-wise smoothness condition (see Remark 2.1 below for details), which has several practical and theoretical drawbacks. Indeed, even for simple volatility processes, it is not

[1]As usual, $a_n \asymp b_n$ if $ma_n \leq b_n \leq Ma_n$, for all n and some $0 < m < M < \infty$.

possible to verify the pathwise Hölder continuity needed for a central limit theorem with the *optimal rate*. Furthermore, even though an 'optimal' bandwidth formula is deduced in closed form therein, this is not well-defined if we want to attain optimal convergence rates for the estimation error (see Remark 2.1 below).

Our contributions. Having discussed some previous work, we now mention some motivating factors and objectives of the present work. To begin with, we wish to impose easily verifiable and general enough conditions to cover a wide range of frameworks without restricting the degree of smoothness of the volatility process. From a theoretical point of view, we also aim to provide a formal justification of the optimal convergence rate of the kernel estimator and to establish central limit theorems (CLT) and asymptotic estimates of the mean square errors with optimal rates. From the practical side, the two factors that affects the performance of the estimator, bandwidth and kernel function, ought to be optimized jointly, not separately, and meanwhile, the proposed method should remain feasible and sufficiently efficient to be implementable for HFD.

The key assumption to our unifying treatment of the problem is a mild local scaling property of the covariance structure of the volatility process. This assumption covers a wide range of frameworks including deterministic differentiable volatility processes and volatilities driven by Brownian Motion, long-memory fractional Brownian Motion, and, more generally, functions of suitable Gaussian processes. Under the referred assumption, we characterize the leading order terms of the Mean Squared Error (MSE) and, as a byproduct, we derive an approximated optimal bandwidth in closed form, which is shown to be asymptotically equivalent to the true optimal bandwidth. From this, the theoretical optimal convergence rate for the estimation error is identified. We then proceed to show that our optimal bandwidth formulas are feasible by proposing an iterated plug-in type algorithm for their implementation. An important intermediate step is to find an estimate of the Integrated Volatility of Volatility (IVV), for which we propose a new estimator based on the two-time scale realized variance of Zhang et al. (2005). Consistency and convergence rate of our vol vol estimator are also established. The estimation of the IVV has also been addressed in Vetter (2015) and Barndorff-Nielsen & Veraart (2009).

Equipped with an explicit formula for the asymptotically optimal MSE, we proceed to set up a well-posed problem for optimal kernel selection. Concretely, for Brownian motion driven volatilities, we prove that the optimal kernel function is the exponential kernel $K(x) = 2^{-1}\exp(-|x|)$. Such a result formalizes and extends a previous result of Foster and Nelson (1994), where only kernels of bounded support were considered. We also show that, due to the nature of the data we are analyzing (namely, HFD), exponential kernel function enjoys outstanding computational advantages, as it reduces the time complexity for estimating the whole path of the volatility on all grid points from $O(n^2)$ to $O(n)$. We also consider the volatility processes driven by the long-memory fractional Brownian motion and, in such a case, we provide numerical schemes to compute the optimal kernel function.

To complement our asymptotic results based on MSE, asymptotic normality of the kernel estimators is also established for two broad types of volatility processes: Itô processes and continuous function of some Gaussian processes. In this way, our results cover volatility processes with flexible degrees of smoothness. The results are consistent with the leading order approximation of the MSE, so that CLT's with the optimal convergence rate are obtained. By contrast, as mentioned above, the CLT's of Fan and Wang (2008) and Kristensen (2010) have suboptimal convergence rate, while the analogous result of Foster and Nelson (1994) is limited to a specific smoothness order and strong constraints on the kernel function and bandwidth. In the case of Itô volatility processes, we generalize the CLT of Alvarez et al. (2012) and Jacod & Protter (2012), from uniform to general forward looking kernels.

Paper Outline. The rest of the paper is organized as follows. In Section 2, we introduce the kernel estimator and our assumptions, and verify that common volatility processes satisfy our assumptions. In Section 3, we deduce the leading order approximation of the MSE of the kernel estimator and solve the optimal bandwidth selection problem. Then, in Section 4, we deal with the optimal kernel function selection problem for different types of volatility processes. A feasible implementation approach of the optimal bandwidth is discussed in Section 5, where we also introduce the two-scale estimator of the IVV. Central Limit Theorems of the kernel estimator are discussed in section 6. Finally in Section 7, we perform Monte Carlo studies. The proofs of the main results are provided in Appendix A while the proofs of some technical lemmas and supporting propositions are deferred to the supplemental article Figueroa-López & Li (2018).

2 Kernel Estimators and Assumptions

In this section, we first introduce the classical kernel estimator for the spot volatility. We then discuss some needed smoothness conditions on the volatility processes and verify that most common volatility processes used in the literature indeed satisfy our assumptions. Finally, we discuss some regularity conditions on the kernel function.

2.1 Framework and Estimators

Throughout the paper, we consider the following stochastic differential equation (SDE):

$$dX_t = \mu_t dt + \sigma_t dB_t, \tag{2.1}$$

where all stochastic processes ($\mu := \{\mu_t\}_{t\geq 0}, \sigma := \{\sigma_t\}_{t\geq 0}, B := \{B_t\}_{t\geq 0}$, etc.) are defined on a complete filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}$. We also assume that μ and σ are adapted to the filtration $\mathbb{F}$ and $B := \{B_t\}_{t\geq 0}$ is a standard Brownian Motion (BM) adapted to $\mathbb{F}$. We suppose throughout the paper that we observe the process X at the times $t_i := t_{i,n} := i\Delta_n$, $0 \leq i \leq n$, where $\Delta_n := T/n$. We will use $\Delta_i^n Z := \Delta Z_{t_{i-1}} := Z_{t_i} - Z_{t_{i-1}}$ to denote the increments of a process Z and $\Delta_n = T/n$ to denote the time increments.

In this paper, we study the problem of estimating the spot volatility σ_τ, at a given time $\tau \in (0, T)$, by the kernel estimator (cf. Fan & Wang (2008) and Kristensen (2010)),

$$\hat{\sigma}^2_{\tau,n,h} := \sum_{i=1}^{n} K_h(t_{i-1} - \tau)(\Delta_i^n X)^2, \tag{2.2}$$

where $K_h(x) = K(x/h)/h$. As is often the case with kernel estimation, the selections of the bandwidth h and kernel function K of (2.2) are of great importance in practice, especially for the finite sample settings commonly encountered in econometric applications. As explained in the introduction, the literature on bandwidth and kernel selection for the spot volatility estimator (2.2) is rather scarce. In this work, we go further with better crafted conditions that allow us to give a unified treatment to the problem for most common volatility processes, including not only Brownian driven volatilities but also those driven by more general Gaussian processes.

2.2 Assumptions on the Volatility Process

Our first assumptions, which is also imposed in Kristensen (2010), is a non-leverage condition. Though not ideal for financial applications, this is imposed for tractability reasons and, in order to allow a unified treatment of volatilities driven by a large spectrum of noises beyond Brownian motion. Our simulations in Section 7 show that this assumption may not be crucial. Indeed, for Brownian-driven volatilities, this assumption is not needed to obtain a central limit theorem for forward-looking kernel functions (see Section 6 for details).

Assumption 1 *(μ, σ) is independent of B.*

Next, we impose some mild moment boundedness assumption on μ and σ.

Assumption 2 *There exists $M_T > 1$ such that $\mathbb{E}[\mu_t^4 + \sigma_t^4] < M_T$, for all $0 \le t \le T$.*

The following is our key assumption, which in the next subsection is shown to be satisfied by a large spectrum of volatility models.

Assumption 3 *Suppose that for $\gamma > 0$ and certain functions $L : \mathbb{R}_+ \to \mathbb{R}_+$ and $C_\gamma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, such that C_γ is not identically zero and*

$$C_\gamma(hr, hs) = h^\gamma C_\gamma(r, s), \quad \text{for } r, s \in \mathbb{R}, h \in \mathbb{R}_+, \tag{2.3}$$

the variance process $V := \{V_t = \sigma_t^2 : t \ge 0\}$ satisfies

$$\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] = L(t) C_\gamma(r, s) + o((r^2 + s^2)^{\gamma/2}), \quad r, s \to 0. \tag{2.4}$$

Remark 2.1 *We now draw some connections with the assumptions and work in Kristensen (2010). Therein, the variance process $\{V_t\}_{t \ge 0}$ is assumed to satisfy the following pathwise condition*

$$|V_{t+\delta} - V_t| \le \tilde{L}(t, |\delta|)|\delta|^\gamma + o(|\delta|^\gamma), \quad \delta \to 0, \tag{2.5}$$

where $\tilde{L}(t, \cdot)$ is a slowly varying random function. To gain some intuition about the plausibility of this assumption, let us suppose that $\{V_t\}$ is a Brownian motion. In that case, the above holds for all $\gamma < 1/2$, but such choices of γ can only produce suboptimal convergence rate of the kernel estimator. On the other, in light of Lévy's modulus of continuity, the condition (2.5) holds for $\gamma = 1/2$, but only if $\tilde{L}(t, \delta) \to \infty$, as $\delta \to 0$. But, in that case, the optimal bandwidth selection formulas obtained in Kristensen (2010) are not well defined as they require that $\lim_{\delta \to 0} \tilde{L}(t, \delta) =: \tilde{L}(t, 0)$ is finite.

A function C_γ satisfying the condition (2.3) is said to be homogeneous of order γ. The index γ determines the degree of smoothness of the volatility paths $t \to \sigma_t$. It is easy to check (see details in the supplemental article Figueroa-López & Li (2018)) that $C_\gamma(r, s; t) := L(t)C_\gamma(r, s)$ is unique and satisfies the following non-negative definiteness property:

$$\iint K(x)K(y)C(x,y)dxdy \geq 0. \tag{2.6}$$

We shall see in the next section that most volatility processes that are studied in the literature satisfy Assumption 3 with a function C_γ of the form:

$$C_\gamma(r,s) = \frac{1}{2}(|r|^\gamma + |s|^\gamma - |r-s|^\gamma), \tag{2.7}$$

for some $\gamma \in [1, 2]$. The case of $\gamma = 1$ covers volatility processes driven by BM, while $\gamma \in (1, 2)$ corresponds to volatility processes driven by fractional Brownian Motions (fBM) with Hurst parameter $H > 1/2$. Deterministic and sufficiently smooth volatility processes can also be incorporated by taking $\gamma = 2$. In the following section, we cover these cases and other more more general models.

2.3 Common Volatility Processes

In this section, we demonstrate that many volatility processes satisfy Assumption 3. We consider three fundamental cases. The proofs of the results in this part are relatively simple and for the sake of space are deferred to the supplemental article Figueroa-López & Li (2018). Let us start by considering the solutions of a standard SDE driven by BM, which are widely used in practice.

Proposition 2.1 *Suppose that the process* $V_t = \sigma^2(t, \omega)$ *satisfies the SDE*

$$dV_t = f(t,\omega)dt + g(t,\omega)dW_t, \quad t \in [0,T], \tag{2.8}$$

where $\{W_t\}_{t\geq 0}$ *is a standard Wiener process adapted to* $\mathbb{F}$*. Assume that* $f(t,\omega)$ *and* $g(t,\omega)$ *are adapted and progressively measurable with respect to* $\mathbb{F}$*,* $\mathbb{E}\left[f^2(t,\omega)\right] < M$*, for* $t \in [0,T]$*, and* $\mathbb{E}\left[g^2(t,\omega)\right]$ *is continuous for* $t \in [0,T]$*. Then, Assumption 3 is satisfied with* $\gamma = 1$*,* $C_1(r,s) = \min\{|r|,|s|\}1_{\{rs\geq 0\}}$*, and* $L(t) = \mathbb{E}[g^2(t,\omega)]$*. Furthermore,* $C_1(r,s)$ *is an integrable positive definite function; i.e., we have strict inequality in (2.6) for all* $K : \mathbb{R} \to \mathbb{R}$ *such that* $\int |K(x)|dx > 0$*.*

In what follows, we show that some processes defined as integrals with respect to a two-sided fBM $B^{(H)} = \{B_t^{(H)} : t \in \mathbb{R}\}$ (see Samorodnitsky & Taqqu (1994) for a detailed survey of fBM) satisfy Assumption 3. A prototypical example is the fractional Ornstein-Uhlenbeck process $Y_t^{(H)} = \sigma \int_{-\infty}^t e^{-\lambda(t-u)} dB_u^{(H)}$, which is frequently used to model volatility processes. It is worth mentioning that, when $H \neq 1/2$, the fBM is not a semimartingale and the problem of defining the stochastic integral with respect to fBM is more subtle. In our paper, we only focus on integrals of deterministic functions f for which the integral can be defined on a path-wise sense under the following condition (cf. Samorodnitsky & Taqqu (1994)):

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(u)f(v)||u-v|^{2H-2}dudv < \infty. \tag{2.9}$$

Proposition 2.2 *Let* $Y_t^{(H)} = \int_{-\infty}^t f(u)dB_u^{(H)}$ *where* $f(\cdot)$ *is a deterministic continuous function that satisfies (2.9) and* $\{B_t^{(H)}\}_{t\in\mathbb{R}}$ *is a (two-sided) fBM with Hurst parameter* $H \in (\frac{1}{2},1)$ *defined on a filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}, P)$*. Then, the processes* $Y^{(H)}$ *and* $\{\exp(Y_t^{(H)})\}_{t\geq 0}$ *satisfy Assumption 3 with* $\gamma = 2H \in (1,2)$ *and* C_γ *given by (2.7).*

For our final case, we show that if a Gaussian process satisfies Assumption 3, so does a suitable smooth function of the process.

Proposition 2.3 *Assume that* $(Z_t)_{t\geq 0}$ *is a Gaussian process that satisfies Assumption 3 uniformly over* $(0,T)$,[2] *with* $\gamma^{(Z)} \in [1,2)$, $L(\cdot)$, *and* $C_\gamma^{(Z)}(\cdot,\cdot)$ *defined as in* (2.4). *For each fixed* $\tau \in (0,T)$ *and a function* $f \in C^2(\mathbb{R})$, *further assume the following:*

(a) $\mathbb{E}[(Z_{\tau+r} - Z_\tau)Z_\tau] = O(|r|)$, $\mathbb{E}[Z_{\tau+r}] - \mathbb{E}[Z_\tau] = O(|r|)$, *as* $r \to 0$.

(b) $\mathbb{E}[(f'(Z_\tau))^4] < \infty$, $\mathbb{E}[\sup_{t\in(\tau-\epsilon,\tau+\epsilon)}(f''(Z_t))^4] < \infty$ *for some* $\epsilon > 0$.

Then, the process $V_t := f(Z_t)$, $t \geq 0$, *satisfies Assumption 3 with* $\gamma^{(V)} = \gamma$ *and* $C_\gamma^{(V)} = \mathbb{E}[(f'(Z_t))^2]C_\gamma^{(Z)}$.

2.4 Conditions on the Kernel

In this part, we introduce the assumptions needed on the kernel function.

Assumption 4 *Given* $\gamma > 0$ *and* C_γ *as defined in Assumption 3, we assume that the kernel function* $K : \mathbb{R} \to \mathbb{R}$ *satisfies the following conditions:*

(1) $\int K(x)dx = 1$;

(2) K *is Lipschitz and piecewise* C^1 *on its support* (A,B), *where* $-\infty \leq A < 0 < B \leq \infty$;

(3) (i) $\int |K(x)||x|^\gamma dx < \infty$; *(ii)* $K(x)x^{\gamma+1} \to 0$, *as* $|x| \to \infty$; *(iii)* $\int |K'(x)|dx < \infty$, *(iv)* $V_{-\infty}^{\infty}(|K'|) < \infty$, *where* $V_{-\infty}^{\infty}(\cdot)$ *is the total variation;*

(4) $\iint K(x)K(y)C_\gamma(x,y)dxdy > 0$.

[2] The Assumption 3 is satisfied uniformly over $(0,T)$ if $\sup_{\tau\in(0,T)}(r^2 + s^2)^{-\gamma/2}\,|\mathbb{E}[(V_{\tau+r} - V_\tau)(V_{\tau+s} - V_\tau)] - L(\tau)C_\gamma(r,s)| \to 0$, as $r,s \to 0$, and, also, $\sup_{\tau\in(0,T)}|L(\tau)| < \infty$. This implies the existence of a positive constant C such that $\mathbb{E}[(Z_t - Z_s)^2] \leq C|t-s|^\gamma$, for all $t,s \in (0,T)$.

3 MSE Decomposition and Bandwidth Selection

In this section, we first deduce an explicit leading order approximation (up to $O(\frac{\Delta}{h})$ and $O(h^\gamma)$ terms) of the MSE of the estimator. The proof is deferred to the Appendix A.

Theorem 3.1 *For the model (2.1) with μ and σ satisfying Assumptions 1-3, and a kernel function K satisfying Assumption 4, let*

$$MSE^a_{\tau,n,h} := 2\frac{\Delta}{h}\mathbb{E}[\sigma^4_\tau]\|K\|^2 + h^\gamma L(\tau)\iint K(x)K(y)C_\gamma(x,y)dxdy. \tag{3.1}$$

Then, for any $\tau \in (0,T)$,

$$MSE_{\tau,n,h} = \mathbb{E}[(\hat{\sigma}^2_\tau - \sigma^2_\tau)^2] = MSE^a_{\tau,n,h} + o\left(\frac{\Delta}{h}\right) + o\left(h^\gamma\right). \tag{3.2}$$

It is not hard to see from the proof of the previous result that all $o(\cdot)$ terms are uniform on $\tau \in (0,T)$ if the condition given by (2.4) is satisfied uniformly in t. Then, we readily get the following:

Corollary 3.1 *Let*

$$MSE^a_{n,h}(a,b) := 2\frac{\Delta}{h}\int_a^b \mathbb{E}[\sigma^4_t]dt\|K\|^2 + h^\gamma \int_a^b L(t)dt \iint K(x)K(y)C_\gamma(x,y)dxdy. \tag{3.3}$$

Then, for the model (2.1) with μ and σ satisfying Assumptions 1-3, so that the term $o((r^2 + s^2)^{\gamma/2})$ in Eq. (2.4) is uniform in t, and a kernel function K satisfying Assumption 4, we have, for any $0 < a < b < T$,

$$IMSE_{n,h} := \int_a^b \mathbb{E}[(\hat{\sigma}^2_t - \sigma^2_t)^2]dt = MSE^a_{n,h}(a,b) + o\left(\frac{\Delta}{h}\right) + o(h^\gamma). \tag{3.4}$$

Based on the approximations above, it is natural to analyze the behavior of the approximated MSE of the kernel estimator. We focus on the integrated MSE (3.4) but an analogous analysis can be made for the local MSE (3.2). Note that, by Assumption 4, we have that $\iint K(x)K(y)C_\gamma(x,y)dxdy > 0$. We then obtain the following:

Proposition 3.1 *With the same assumptions as Corollary 3.1, the approximated optimal homogeneous bandwidth, denoted by* $\bar{h}_n^{a,opt}$*, which is defined to minimize the approximated IMSE given by (3.3), is given by*

$$\bar{h}_n^{a,opt} = n^{-1/(\gamma+1)} \left[\frac{2T \int_a^b \mathbb{E}[\sigma_t^4]dt \int K^2(x)dx}{\gamma \int_a^b L(t)dt \iint K(x)K(y)C_\gamma(x,y)dxdy} \right]^{1/(\gamma+1)}, \tag{3.5}$$

while the attained minimum of the approximated IMSE is given by

$$\begin{aligned} IMSE_n^{a,opt}(a,b) = n^{-\gamma/(1+\gamma)} \left(1 + \frac{1}{\gamma}\right) \left(2T \int_a^b \mathbb{E}[\sigma_t^4]dt \int K^2(x)dx\right)^{\gamma/(1+\gamma)} \\ \times \left(\gamma \int_a^b L(t)dt \iint K(x)K(y)C_\gamma(x,y)dxdy\right)^{1/(1+\gamma)}. \end{aligned} \tag{3.6}$$

A direct consequence of the previous result is the following proposition about the optimal convergence rate.

Proposition 3.2 *With the same assumptions as those in Corollary 3.1, the optimal convergence rate of the kernel estimator is given by* $n^{-\gamma/(1+\gamma)}$*. This is attainable if the bandwidth is chosen as* $h_n = cn^{-1/(\gamma+1)}$ *for some constant* $c \in (0, \infty)$*.*

An important problem is to formalize the connection between the approximate optimal bandwidth $\bar{h}_n^{a,opt}$ (respectively, $h_n^{a,opt}$), which is defined as the minimizer of the MSE (3.3) (respectively, (3.1)), and the "true" optimal bandwidth, whenever it exists, which is denoted by $\bar{h}_n^*$ (respectively, h_n^*) and is defined as a value of the bandwidth that minimizes the actual IMSE (respectively, MSE) of the kernel estimator. In the supplemental article Figueroa-López & Li (2018), we show that, under a mild additional condition, they are equivalent in the sense that $\bar{h}_n^* = \bar{h}_n^{a,opt} + o(\bar{h}_n^{a,opt})$ and $h_n^* = h_n^{a,opt} + o(h_n^{a,opt})$.

4 Kernel Function Selection

As an important application of the optimal bandwidth selection problem defined in Section 3, we now study the problem of selecting an optimal kernel function by minimizing the optimal IMSE attained by (3.5). As shown therein, the optimal kernel function only depends on the covariance structure, $C_\gamma(\cdot,\cdot)$. There are two possible situations. The first one is when C_γ is positive definite. In such a case, we cannot improve the rate of convergence of the IMSE, but we can attempt to minimize the constant appearing before the asymptotics of the IMSE in (3.6) or, equivalently, minimize the functional:

$$I(K) = \left(\int K^2(x)dx\right)^\gamma \int\int K(x)K(y)C_\gamma(x,y)dxdy. \tag{4.1}$$

Another situation is when C_γ is simply non-negative definite. In such a case, if we relax (4) of Assumption 4, it is possible to improve the rate of convergence of the IMSE by choosing a so-called "higher order" kernel function. An important instance of this case is when the volatility is deterministic and sufficiently smooth (see Remark 4.1 below for more information).

In this section, we focus on the covariance function C_γ defined in Eq. (2.7) with $\gamma < 2$, which is actually positive definite. This is because C_γ admits the integral form $C_\gamma(x,y) = \int F_\gamma(x,u)F_\gamma(y,u)du$ with

$$F_\gamma(x,y) := m\left(|x-y|^{\frac{\gamma-1}{2}}\mathrm{sgn}(x-y) + |y|^{\frac{\gamma-1}{2}}\mathrm{sgn}(y)\right),$$

and a certain constant m (see Ossiander & Waymire (1989) for details). We can then easily check that $\int\int K(x)K(y)C_\gamma(x,y)dx = \int(\int K(x)F_\gamma(x,u)dx)^2du > 0$, for an arbitrary nonzero kernel function $K : \mathbb{R} \to \mathbb{R}$. Furthermore, it also follows that its symmetrization, $K_s(x) := (K(x) + K(-x))/2$, is such that

$$\iint K(x)K(y)C_\gamma(x,y)dxdy - \iint K_s(x)K_s(y)C_\gamma(x,y)dxdy \geq 0. \tag{4.2}$$

The previous relation implies that in order to minimize the constant appearing before the asymptotic IMSE in (3.5), it suffices to consider symmetric kernel functions K.

Remark 4.1 *In the accompanying article Figueroa-López & Li (2018), we give some new results regarding optimal kernel selection for smooth deterministic volatilities. Concretely, by using the calculus of variation with constraints, we obtain optimal kernel functions of higher orders. The second order optimal kernel is exactly that of Epanechnikov (1969) kernel, while, for higher order cases, we provide ways to calculate those optimal kernel functions.*

4.1 Optimal Kernel Selection for BM driven Volatilities

Consider a BM type volatility with $\gamma = 1$ and $C_1(r, s) = 1_{\{rs>0\}} \min(|r|, |s|)$. We will show that the exponential kernel function is the optimal kernel function. Foster and Nelson (1994) argued that this is the case, but their proof lacks rigor, due to their bounded support assumption on the kernel function.

From (4.1) and the relation (4.2), the objective function that we need to minimize is

$$\int_0^\infty K^2(x)dx \int_0^\infty \int_0^\infty K(x)K(y)\min(x,y)dxdy = \int_0^\infty [L'(x)]^2 dx \int_0^\infty [L(x)]^2 dx =: I^*(L),$$

where we set $L(u) := \int_u^\infty K(x)dx$. The problem is then changed to minimize $I^*(L)$ for functions L that are continuous and piece-wise twice differentiable on $\mathbb{R}_+$ such that $L(0) = \frac{1}{2}$ and $\lim_{x\to+\infty} L(x) = 0$. Next, using Cauchy-Schwartz inequality, note that

$$I^*(L) \geq \left(\int_0^\infty L'(x)L(x)dx\right)^2 = \left(\int_0^\infty L(x)dL(x)\right)^2 = \left(\int_{1/2}^0 udu\right)^2 = \frac{1}{64},$$

where the first inequality becomes equality if and only if there exist non-zero constants C_1 and C_2 such that $C_1 L'(x) + C_2 L(x) \equiv 0$, for all $x \in \mathbb{R}_+$. We have two possible cases: (1)

there exists $x_0 > 0$, such that $L(x) > 0$, for all $x \in [0, x_0)$ and $L(x_0) = 0$; (2) $L(x) > 0$, for all $x \in \mathbb{R}_+$. For the first case, we have that $L'(x)/L(x) = -C_2/C_1$, for $x \in (0, x_0)$, whose solution is $L(x) = \frac{1}{2}e^{Bx}$ and it is then impossible that $L(x_0) = 0$. Therefore, only the second case is possible and, by solving the same differential equation, we have the following.

Theorem 4.1 *For the model (2.1) with μ and σ satisfying Assumptions 1-3, where C_γ is given by (2.7) with $\gamma = 1$, and for a kernel function K satisfying Assumption 4, we have that the optimal kernel function that minimizes the first order approximation of the IMSE of the kernel estimator is the exponential kernel function $K^{exp}(x) = \frac{1}{2}\exp(-|x|)$.*

Remark 4.2 *We can easily demonstrate to what extent the exponential kernel decreases the MSE. As seen from (3.6), $IMSE_n^{a,opt} = C\sqrt{I^*(K)}$, where the constant C does not depend on the kernel function K. Below, we show the value of $I^*(K) := I^*(L)$ for the exponential, uniform, triangular, and the Epanechnikov kernels:*

$$I^*(.5\, e^{-|x|}) = \frac{1}{72} \approx 0.0138, \quad I^*(.5\, 1_{\{|x|<1\}}) = \frac{1}{24} \approx 0.0416,$$
$$I^*(|1-x|1_{\{|x|<1\}}) = \frac{1}{30} \approx 0.0333, \quad I^*(.75\,(1-x^2)1_{\{|x|<1\}}) = \frac{297}{8240} \approx 0.036.$$

Let us finish by noting that the exponential kernel function not only minimizes the MSE of the kernel estimator, but also enables us to substantially reduce the computational complexity of the volatility estimation. The idea is to use the following recurrent algorithm:

$$\begin{aligned}
\hat{\sigma}^2_{\tau+\Delta,-} &= e^{-\Delta/h}\left[\hat{\sigma}^2_{\tau,-} + K_h^{exp}(t_{i_0-1}-\tau)(\Delta_{i_0}X)^2\right], \\
\hat{\sigma}^2_{\tau+\Delta,*} &= K_h^{exp}(t_{i_0}-(\tau+\Delta))(\Delta_{i_0+1}X)^2, \\
\hat{\sigma}^2_{\tau+\Delta,+} &= e^{\Delta/h}\left[\hat{\sigma}^2_{\tau,+} - K_h^{exp}(t_{i_0}-\tau)(\Delta_{i_0+1}X)^2\right],
\end{aligned} \tag{4.3}$$

where i_0 is such that $t_{i_0-1} \le \tau < t_{i_0}$, $K_h^{exp}(x) = \frac{1}{2h}\exp(-|x|/h)$, and

$$\hat{\sigma}^2_{\tau,-} = \sum_{i<i_0} K_h^{exp}(t_{i-1}-\tau)(\Delta_i X)^2, \quad \hat{\sigma}^2_{\tau,+} = \sum_{i>i_0} K_h^{exp}(t_{i-1}-\tau)(\Delta_i X)^2.$$

It is now clear that, in order to estimate $\{\sigma_{t_i}\}_{i=0,\ldots,n}$, using an exponential kernel, we need a time of $O(n)$, instead of the orders $O(n^2)$ or $O(n^2 h)$ needed for general kernels of unbounded or bounded support, respectively.

In practice, kernel estimators suffer of biases at times closer to the boundary. As proposed in Kristensen (2010), this can be corrected by using the following estimator:

$$\hat{\sigma}^b_{\tau,n,h} = \frac{\sum_{i=1}^n K_h(t_{i-1}-\tau)(\Delta_i^n X)^2}{\Delta \sum_{i=1}^n K_h(t_{i-1}-\tau)}. \tag{4.4}$$

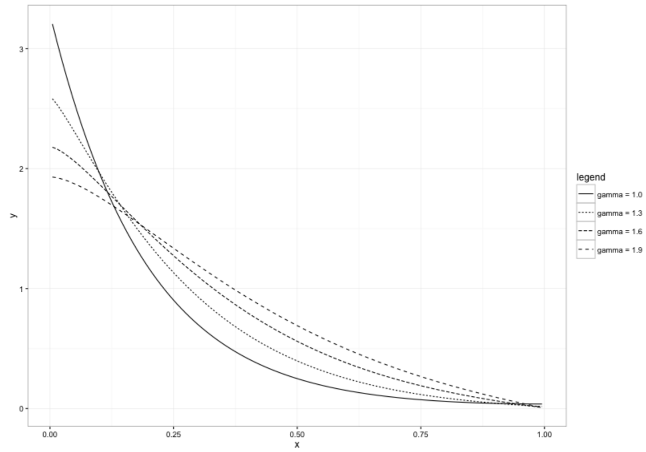
where the superscript denotes boundary effect. The denominator above can still be efficiently calculated similarly as (4.3) except that all $(\Delta_i X)^2$ are replaced by 1.

4.2 Optimal Kernel Function for a fBM driven Volatility

In this section, we now consider a general fBM covariance structure, i.e. $\gamma \in (1,2)$ and C_γ given by (2.7). Again, our goal is to minimize (4.1). As implied by the relation (4.2), we only need to consider symmetric kernel functions. Unfortunately, the problem of finding an explicit form for the optimal kernel function is more challenging. Therefore, we instead seek for a numerical method to find the optimal kernel function, for which, we consider two approximation. First, since all unbounded support kernels can be approximated by a kernel with a bounded support and the optimization problem is unchanged with $K(x)$ scaled by a small bandwidth, we will limit the support of $K(x)$ to be $[0,1]$. Second, we approximate the kernel function K by step functions of the form $K_m(x) = \frac{1}{\sum_{i=1}^n a_i}\sum_{i=1}^m a_i 1_{[\frac{i-1}{m},\frac{i}{m})}(|x|)$, with $x \in [-1,1]$, $a_i \in \mathbb{R}$ $(i=1,\ldots,n)$, and then use gradient descent to directly optimize (4.1) over all valid values of $(a_1,\ldots,a_m)$. In spite of the high dimensionality of the optimization problem, this is still tractable, since the gradient can be calculated explicitly.

Figure 1 shows the resulting optimal kernels for $\gamma = 1.0, 1.3, 1.6, 1.9$. Note that the resulting approximated optimal kernel for $\gamma = 1$ is consistent with true optimal kernel that was proved to be exponential in Section 4.1. We also observe from Figure 1 that, as γ increases, the optimal kernel function becomes flatter and less convex. This indeed makes sense, since a higher γ indicates less chaos of the volatility, and thus more weights should be given to data farther from the estimated point.

Figure 1: Optimal Kernel Functions for Different γ



5 Plug-In Bandwidth Selection Methods

In this section we propose a feasible plug-in type bandwidth selection algorithm, for which, as a sub-problem, we also develop a new estimator of the volatility of volatility based on the kernel estimator of the spot volatility and a type of two-time scale realized variance estimator. We shall focus on the case of a BM type volatility as described in Proposition 2.1, while similar methods can be developed for other types of volatility structures. To implement the approximated optimal bandwidth formula (3.5), it is natural to estimate $\int_0^T \mathbb{E}[\sigma_t^4]dt$ and $\int_a^b L(t)dt = \int_a^b \mathbb{E}[g^2(t)]dt$ with the integrated quarticity of X, $IQ(X) = \int_0^T \sigma_\tau^4 d\tau$, and the quadratic variation of σ^2, $IV(\sigma^2) = \int_0^T g^2(\tau)d\tau$. A popular estimate for $\int_0^T \sigma_\tau^4 d\tau$ is the realized quarticity, which is defined by $\widehat{IQ} = (3\Delta)^{-1} \sum_{i=1}^{n} (\Delta_i X)^4$. The estimation of $\int_0^T g^2(\tau)d\tau$ is a more subtle problem and, below, we propose an estimator, which is termed the Two-time Scale Realized Volatility of Volatility (TSRVV) and is hereafter denoted by $\widehat{IV(\sigma^2)}_{(\text{tsrvv})}$. With these estimators, the final bandwidth can then be written as

$$h_n^{a,opt} = \left[\frac{2T\widehat{IQ(X)} \int K^2(x)dx}{n\widehat{IV(\sigma^2)}_{(\text{tsrvv})} \iint K(x)K(y)C_1(x,y)dxdy} \right]^{1/2}. \tag{5.1}$$

The previous bandwidth estimator involves the spot volatility itself, through $\widehat{IV(\sigma^2)}_{(\text{tsrvv})}$, which, of course, we do not know in advance. To deal with this problem, we propose to use an iterative algorithm in the same spirit of a fixed-point type of procedure. Concretely, we start with an initial "guess" for the bandwidth such as

$$h_n^{init} = \left[\frac{2T \int K^2(x)dx}{n \iint K(x)K(y)C_1(x,y)dxdy} \right]^{1/2}. \tag{5.2}$$

With such a bandwidth, we can obtain initial estimates of the spot volatility at all the grid points. Such an initial spot volatility estimation can then be applied to compute $\widehat{IV(\sigma^2)}_{(\text{tsrvv})}$, which, in turn, can be used to obtain another estimation of the optimal bandwidth. This procedure is continued iteratively until a predetermined stopping criteria is met. Our simulations show that one or two iterations are typically enough.

We are now ready to define our estimator $\widehat{IV(\sigma^2)}_{(\text{tsrvv})}$ of $IV(\sigma^2) = \int_0^T g^2(\tau)d\tau$, which is often referred to as the Integrated Volatility of Volatility (IVV) of X. The idea is to note that, at each observation time t_i, the estimated spot volatility can be written as $\hat{\sigma}_{t_i}^2 = \sigma_{t_i}^2 + e_{t_i}$, where e_{t_i} is the estimation error. This suggests to make an analogy with the problem of estimating the realized quadratic variation of a semimartingale Y based on discrete observations of Y exposed to market microstructure. So, we can apply any of the different techniques to tackle this problem such as the Two-time Scale Realized Volatility

(TSRV) estimator of Zhang et al. (2005). However, note that, unlike the problem in Zhang et al. (2005), our estimation errors are correlated and such a correlation becomes more significant when we take the difference $\Delta_i \hat{\sigma}^2 = \hat{\sigma}^2_{t_{i+1}} - \hat{\sigma}^2_{t_i}$. To alleviate such a problem, we propose to use one-sided kernel estimators and take the difference between the right and left side estimators to find $\Delta_i \hat{\sigma}^2$. Concretely, let $\hat{\sigma}^2_{l,t_i}$ and $\hat{\sigma}^2_{r,t_i}$ be the left and right side estimator of $\sigma^2_{t_i}$, respectively, defined as

$$\hat{\sigma}^2_{l,t_i} = \frac{\sum_{j>i} K_h(t_{j-1} - t_i)(\Delta_j^n X)^2}{\Delta \sum_{j>i} K_h(t_{j-1} - \tau)}, \quad \hat{\sigma}^2_{r,t_i} = \frac{\sum_{j\leq i} K_h(t_{j-1} - t_i)(\Delta_j^n X)^2}{\Delta \sum_{j\leq i} K_h(t_{j-1} - \tau)}. \tag{5.3}$$

Next, we define the following two difference terms: $\Delta_i \hat{\sigma}^2 = \hat{\sigma}^2_{r,t_{i+1}} - \hat{\sigma}^2_{l,t_i}$, $\Delta_i^{(k)} \hat{\sigma}^2 = \hat{\sigma}^2_{r,t_{i+k}} - \hat{\sigma}^2_{l,t_i}$. Finally, we can construct the estimator

$$\widehat{IVV}_{(\text{tsrvv})} = \frac{1}{k} \sum_{i=b}^{n-k-b} (\Delta_i^{(k)} \hat{\sigma}^2)^2 - \frac{n-k+1}{nk} \sum_{i=b+k-1}^{n-k-b} (\Delta_i \hat{\sigma}^2)^2. \tag{5.4}$$

Here, b is a small enough integer, when compared to n. The purpose of introducing such a number b is to alleviate the boundary effect of the one sided estimators. Similar to Zhang et al. (2005), we can take $k = n^{2/3}$ in our case. There is some work to do if one wants to optimize such a TSRVV estimator, but this is outside the scope of the present work.

The result below shows the consistency of (5.4) and shed some light on its rate of convergence. Its proof cis given in the supplemental article Figueroa-López & Li (2018).

Theorem 5.1 *Fix a $t_b \in (0, T/2)$. Then, for the model (2.1) with μ and σ satisfying Assumptions 1 and 2 and σ being a squared integrable Itô process as in Eq.* (2.8) *(thus satisfying Assumption 3), and a kernel function K satisfying Assumption 4,* (5.4) *is a consistent estimator of $\int_{t_b}^{T-t_b} g_t^2 dt$ with $b = t_b/\Delta$. Furthermore, the convergence rate is given by $O_p(\frac{n^{1/4}}{k^{1/2}}) + O_p(\sqrt{\frac{k}{n}})$.*

Remark 5.1 *Vetter (2015) proposed a similar estimator for the IVV, but taking a right-sided* uniform *kernel when computing the difference $\Delta_i \hat{\sigma}^2$ of the estimated volatility and also applying a different bias correction technique from ours. It is shown therein that his estimator attains the optimal rate of convergence of $n^{-1/4}$. Simulations, that are not shown here for the sake of space, indicate that our TSRVV using the optimal exponential kernel has better performance than Vetter (2015) at least for the chosen parameter choices. This suggests that there may be some room for improvement of the convergence rate stated in Theorem 5.1. On the other hand, the observed improved performance of our TSRVV may be a consequence of the fact that we are using an exponential kernel, while the estimator in Vetter (2015) uses the suboptimal uniform kernel.*

6 Central Limit Theorems

In this section, we aim to characterize the limiting distribution of the estimation error of the kernel estimator by proving a Central Limit Theorem (CLT). To motivate the discussion below, let us start by noting the following natural decomposition:

$$\begin{aligned}\hat{\sigma}^2_\tau - \sigma^2_\tau = &\sum_{i=1}^n K_h(t_{i-1}-\tau)(\Delta_i X)^2 - \int_0^T K_h(t-\tau)\sigma_t^2 dt \\ &+ \int_0^T K_h(t-\tau)(\sigma_t^2 - \sigma_\tau^2)dt + o_p\left(h^\gamma\right),\end{aligned} \tag{6.1}$$

where the last term on the right-hand side above follows from Assumption 4. Two general type of results can be found in the literature to deal with the estimation error:

(1) One approach consists of using a 'suboptimal' bandwidth so that the first error term in (6.1), which, as shown below, is of order $O_p((\Delta/h)^{1/2})$, dominates the second term, whose order is $O_p(h^{\gamma/2})$. This would be the case if, for instance, we choose

$h = o(\Delta^{1/(\gamma+1)})$. Instances of this type of results can be found in Fan & Wang (2008) and Kristensen (2010).

(2) In the case that σ_t^2 follows an Itô process, Foster & Nelson (1994) obtained a CLT for the kernel estimator $\hat{\sigma}_\tau^2$ with optimal convergence rate but under a number of stringent conditions. In particular, only kernels with bounded support were considered. More recently, under relatively mild assumptions in the Itô dynamics of X and σ, Alvarez et al. (2012) and Jacod & Protter (2012) also obtained a CLT with optimal convergence rate but only for a forward-looking uniform kernel function.

The two previous approaches have some obvious limitations. The first approach can only yield results with suboptimal convergence rates, while the second type of results only deal with one level of smoothness in the volatility process. In this section, we obtain a CLT for two broad frameworks: (i) Itô type volatilities and (ii) deterministic functions of certain Gaussian processes. These cover all the examples mentioned in Section 2.3. For the framework (i), we consider two cases: 1) A general kernel but no leverage; 2) Leverage but only forward looking kernel as in Alvarez et al. (2012) and Jacod & Protter (2012), even though these two works only consider uniform kernels, while we consider here a general forward-looking kernel function. The second framework (ii) covers a wide range of models of different smoothness order, though without leverage. In what follows, we replace Assumption 1 with the following:

Assumption 5 *The processes μ and σ are adapted cádlág.*

We begin with an analysis of the first error term in (6.1), which, in the nonleverage case was already studied in Kristensen (2010). Its proof uses the CLT for martingale differences and is similar to that of Theorem 2.7 in Mancini et al. (2004), but, since the proof in

Mancini et al. (2004) is for a more general class of estimators and requires more technical analysis, we give a simpler proof in the supplemental article Figueroa-López & Li (2018).

Theorem 6.1 *For the model (2.1) with μ and σ satisfying Assumption 5, and a kernel function K satisfying Assumption 4, we have, for any $\tau \in (0, T)$,*

$$\left(\frac{\Delta}{h}\right)^{-1/2} \left[\sum_{i=1}^{n} K_h(t_{i-1} - \tau)(\Delta_i X)^2 - \int_0^T K_h(t - \tau)\sigma_t^2 dt\right] \to_D \delta_1 N(0, 1), \tag{6.2}$$

where $\delta_1^2 = 2\sigma_\tau^4 \int K^2(x)dx$.

Next, we consider the second error term in (6.1), which only involves properties of the volatility process σ and not the interaction between X and σ. The proof is provided in Appendix A.

Theorem 6.2 *Let K be a kernel function satisfying Assumption 4 and fix a $\tau \in (0, T)$. Additionally, suppose that either one of the following conditions holds:*

(1) $\{\sigma_t^2\}_{t\geq 0}$ is an Itô process given by $\sigma_t^2 = \sigma_0^2 + \int_0^t f_s ds + \int_0^t g_s dW_s$ with adapted cádlág processes $\{f_t\}_{t\geq 0}$ and $\{g_t\}_{t\geq 0}$.

(2) $\sigma_t^2 := f(Z_t)$, $t \in [0, T]$, for a deterministic function $f : \mathbb{R} \to \mathbb{R}$ and a Gaussian process $\{Z_t\}_{t\geq 0}$ satisfying all requirements of Proposition 2.3.

Then, on an extension $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{\mathbb{P}})$ of the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, equipped with a standard normal variable ξ independent of g_τ in (1) or Z_τ in (2), we have:

$$h^{-\gamma/2}\left(\int_0^T K_h(t - \tau)(\sigma_t^2 - \sigma_\tau^2)dt\right) \to_D \delta_2 \xi, \tag{6.3}$$

where, under the condition (1) above, $\delta_2^2 = g_\tau^2 \iint K(x)K(y)C(x, y)dxdy$, while, under the condition (2), $\delta_2^2 = [f'(Z_\tau)]^2 L^{(Z)}(\tau) \iint K(x)K(y)C_\gamma^{(Z)}(x, y)dxdy$.

As a byproduct of Theorems 6.1 and 6.2 and in accordance with our former Proposition 3.2, we deduce that the optimal convergence rate is $n^{-\gamma/(1+\gamma)}$ and that this would be attained if $h_n = c\Delta_n^{1/(\gamma+1)}$ for any constant $c \in (0, \infty)$. In that case, the following result shows a CLT for $\hat{\sigma}_\tau^2$ under the non-leverage Assumption 1 (its proof is omitted for the sake of brevity and presented in the supplemental article Figueroa-López & Li (2018):

Corollary 6.1 *Suppose the assumptions of Theorems 6.1 and 6.2 are satisfied as well as the nonleverage Assumption 1. Then, for the optimal bandwidth* $h_n = \Delta^{1/(\gamma+1)}$*, we have* $\Delta^{-\frac{\gamma}{2(1+\gamma)}}(\hat{\sigma}_\tau^2 - \sigma_\tau^2) \to_D \sqrt{\delta_1^2 + \delta_2^2}\,\bar{\xi}$*, where* δ_1 *and* δ_2 *are defined in Theorems 6.1 and 6.2, respectively, and* $\bar{\xi}$ *is a standard normal random variable independent from* g_t*, under the setting (1) of Theorem 6.2, or from* Z_t *under the setting (2) of Theorem 6.2.*

Our final result shows a CLT for $\hat{\sigma}_\tau^2$ when $h_n = cn^{-1/(\gamma+1)}$ for general Itô volatilities (as in the setting (1) of Theorem 6.2), but only forward looking kernels. This generalizes results of Alvarez et al. (2012) and Jacod & Protter (2012), where only uniform forward kernels were considered. Its proof is given in the supplemental article Figueroa-López & Li (2018).

Theorem 6.3 *Consider the model (2.1) with a cádlag process* μ *and an Itô process* σ *given by* $\sigma_t^2 = \sigma_0^2 + \int_0^t f_s ds + \int_0^t g_s dW_s$ *where* W *is a Brownian motion such that* $\mathbb{E}(dB_t \cdot dW_t) = \rho dt$ *and* $\{f_t\}_{t\geq 0}$ *and* $\{g_t\}_{t\geq 0}$ *are adapted cádlág processes. Let* K *be a kernel function satisfying Assumption 4 and, in addition,* $K(x) = 0$ *for all* $x < 0$*. Then, the conclusion of Corollary 6.1 holds true with* $\gamma = 1$*.*

7 Simulation Results

In this section, we show some simulations to further investigate the performance of the plug-in method that we developed in Sections 3 and 5 and compare it with the cross-validation method proposed in Kristensen (2010). Throughout, we will consider the Heston model:

$$dX_t = \mu_t dt + \sqrt{V_t} dB_t, \quad dV_t = \kappa(\theta - V_t)dt + \xi\sqrt{V_t} dW_t. \tag{7.1}$$

As to the parameters values, we adopt the setting used in Zhang et al. (2005):

$$\kappa = 5, \quad \theta = 0.04, \quad \xi = 0.5, \quad \mu_t = 0.05 - V_t/2.$$

The initial values are set to be $X_0 = 1, \sigma_0^2 = 0.04$. We also assume both a non-leverage setting ($\rho = 0$) and a negative leverage situation ($\rho = -0.5$) to investigate the robustness of our method against non-zero ρ values. We will consider several different sampling scenarios with 6.5 trading hours per day (the time unit is one year).

In order to alleviate boundary effect, we use the estimator (4.4) throughout all the simulation. For each simulated discrete skeleton $\{X_{t_i} : 0 \le i \le n, t_i = iT/n\}$, we estimate the corresponding discrete-skeleton of the variance process $\{\sigma_{t_i}^2 : 0 \le i \le n\}$, and calculate the average of the squared errors, $ASE = \frac{1}{n-2l+1}\sum_{i=l}^{n-l}(\hat{\sigma}_{t_i}^2 - \sigma_{t_i}^2)^2$, for each simulation. We use $l = [0.1n]$ to focus on evaluating the performance of the estimator without boundary effects. Then, we take the sample average of such ASE's to estimate the mean ASE, defined as $\mathrm{MASE} = \mathbb{E}\left[\frac{1}{n-2l+1}\sum_{i=l}^{n-l}(\hat{\sigma}_{t_i}^2 - \sigma_{t_i}^2)^2\right]$.

In Table 1, we report the MASE obtained by different methods based on 2000 paths. The first column is the plug-in method, where we use the approximated homogeneous optimal bandwidth (3.5) with parameters estimated as proposed in Section 5. In the second column, we report the result for the leave-one-out cross validation as proposed in Kristensen (2010). In the third column, we report the result for an oracle plug-in method, where the true σ and ξ are used to compute $\int_0^T \sigma_\tau^4 d\tau$ and $\int_0^T g^2(\tau)d\tau = \xi^2\int_0^T \sigma_\tau^2 d\tau$ in the formula (3.5). The final column shows a "semi-oracle" result, which assumes the knowledge of the volatility of volatility ξ of the Heston model, but not σ.

As expected, the plug-in method runs significantly faster than cross validation. As to the accuracy of the kernel estimator, simulation results show that, in almost all sampling frequencies, the plug-in method outperforms the cross-validation method. It is worth to notice that, in all cases, there is still significant loss of accuracy for the plug-in method compared to the oracle ones. From the two oracle results, it can be easily observed that such a loss of accuracy is mainly due to the estimation error of the volatility of volatility. Further investigation of the estimation of the volatility of volatility is an interesting and important topic for future research.

5 Days Data

nData/h	ρ	$MASE_{PI}$	$MASE_{CV}$	$MASE_{oracle}$	$MASE_{semi-oracle}$
12	0	1.0796E-07	1.3386E-07	9.1266E-08	9.0402E-08
60	0	7.1439E-09	8.0542E-09	6.7286E-09	6.7074E-09
12	-0.5	1.0296E-07	1.4180E-07	9.2620E-08	9.2009E-08
60	-0.5	7.3872E-09	8.2567E-09	6.9356E-09	6.9060E-09

21 Days Data

nData/h	ρ	$MASE_{PI}$	$MASE_{CV}$	$MASE_{oracle}$	$MASE_{semi-oracle}$
12	0	1.9088E-08	2.1221E-08	1.8265E-08	1.8178E-08
60	0	1.7064E-09	1.6868E-09	1.5984E-09	1.5961E-09
12	-0.5	1.9039E-08	1.9495E-08	1.7587E-08	1.7506E-08
60	-0.5	1.6652E-09	1.6011E-09	1.5509E-09	1.5505E-09

Table 1: Comparison of Different Bandwidth Selection Methods (MASE, 2000 simulations)

A Proofs of Main Results

Proof of Theorem 3.1. Let us start by writing the MSE as follows:

$$\text{MSE} = \mathbb{E}\left[\left(\sum_{i=1}^{n} K_h(t_{i-1}-\tau)((\Delta_i X)^2 - \Delta\sigma_\tau^2) + \left(\sum_{i=1}^{n} K_h(t_{i-1}-\tau)\Delta - 1\right)\sigma_\tau^2\right)^2\right].$$

By Lemmas 3.1 and 3.2 of the supplemental article Figueroa-López & Li (2018), with $f(t) \equiv 1$, we have $\sum_{i=1}^n K_h(t_{i-1}-\tau)\Delta - 1 = O(\Delta/h) + o(h^\gamma)$ and, thus,

$$\text{MSE} = \sum_{i,j=1}^{n} K_h(t_{i-1}-\tau)K_h(t_{j-1}-\tau)\mathbb{E}[((\Delta_i X)^2 - \Delta\sigma_\tau^2)((\Delta_j X)^2 - \Delta\sigma_\tau^2)] + \text{h.o.t.},$$

where hereafter h.o.t. refers to terms of order $o(\Delta/h) + o(h^\gamma)$. Next, applying Lemmas 3.1 and 3.2 of the supplemental article Figueroa-López & Li (2018), together with Assumptions 1 and 2, we have

$$\begin{aligned} MSE = &\sum_{i,j=1}^{n} K_h(t_{i-1}-\tau)K_h(t_{j-1}-\tau) \\ &\times \mathbb{E}\left[\left(\left(\int_{t_{i-1}}^{t_i}\sigma_t dB_t\right)^2 - \Delta\sigma_\tau^2\right)\left(\left(\int_{t_{j-1}}^{t_j}\sigma_t dB_t\right)^2 - \Delta\sigma_\tau^2\right)\right] + \text{h.o.t.}. \end{aligned}$$

Next, by Assumption 1, it follows that

$$\begin{aligned} MSE = &\,2\sum_{i=1}^{n} K_h^2(t_{i-1}-\tau)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i}\sigma_t^2 dt\right)^2\right] \\ &+ \sum_{i,j=1}^{n} K_h(t_{i-1}-\tau)K_h(t_{j-1}-\tau)\int_{t_{i-1}}^{t_i}\int_{t_{j-1}}^{t_j}\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)(\sigma_s^2-\sigma_\tau^2)]dtds + \text{h.o.t.} \\ &=: 2V_1 + V_2 + \text{h.o.t.}. \end{aligned}$$

We now proceed to analyze V_1 and V_2. Firstly, for V_1, note that

$$\begin{aligned}\mathbb{E}\left(\int_{t_{i-1}}^{t_i}\sigma_t^2 dt\right)^2 &= \Delta^2\mathbb{E}[\sigma_\tau^4] + 2\Delta\int_{t_{i-1}}^{t_i}\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)\sigma_\tau^2]dt + \mathbb{E}\left(\int_{t_{i-1}}^{t_i}(\sigma_t^2-\sigma_\tau^2)dt\right)^2\\ &=: \Delta^2\mathbb{E}[\sigma_\tau^4] + B_i + C_i.\end{aligned}$$

To analyze the contribution of each of the three terms above to V_1, we use Lemmas 3.1 and 3.2 of the supplemental article Figueroa-López & Li (2018) with 'kernel' function K^2 and the following three different functions f:

$$f(t) = 1, \quad f(t) = \sqrt{\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2]\mathbb{E}[\sigma_\tau^4]}, \quad f(t) = \mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2],$$

respectively. It then follows that

$$\begin{aligned}&\Delta^2\sum_{i=1}^n K_h^2(t_{i-1}-\tau) = \frac{\Delta}{h}\sum_{i=1}^n K^2(\frac{t_{i-1}-\tau}{h})\frac{\Delta}{h} = \frac{\Delta}{h}\int K^2(x)dx + \text{h.o.t.},\\ &\sum_{i=1}^n K_h^2(t_{i-1}-\tau)B_i \leq 2\frac{\Delta}{h}\sum_{i=1}^n K^2(\frac{t_{i-1}-\tau}{h})\frac{1}{h}\int_{t_{i-1}}^{t_i}\sqrt{\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2]\mathbb{E}[\sigma_\tau^4]}dt = \text{h.o.t.},\\ &\sum_{i=1}^n K_h^2(t_{i-1}-\tau)C_i \leq \frac{\Delta}{h}\sum_{i=1}^n K^2(\frac{t_{i-1}-\tau}{h})\frac{1}{h}\int_{t_{i-1}}^{t_i}\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2]dt = \text{h.o.t.},\end{aligned}$$

where the second line above follows from the fact that $\mathbb{E}[(\sigma_t^2-\sigma_\tau^2)^2] = O(|t-\tau|^\gamma)$. Putting together the previous relationships, we conclude that

$$V_1 = \sum_{i=1}^n K_h^2(t_{i-1}-\tau)\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i}\sigma_t^2 dt\right)^2\right] = \frac{\Delta}{h}\mathbb{E}[\sigma_\tau^4]\int K^2(x)dx + \text{h.o.t.}.$$

Next, applying directly Lemmas 3.1 and 3.2 of the supplemental article Figueroa-López & Li (2018)together with Assumption 3, V_2 can be written as

$$V_2 = h^\gamma\int\int K(x)K(y)C_\gamma(x,y;\tau)dxdy + o\left(\frac{\Delta}{h}\right) + o(h^\gamma).$$

■

Proof of Theorem 6.2. **(1)** As it is standard in the literature, by virtue of localization (as in Jacod and Shiryaev, section 5.4, p.549), we assume without loss of generality that the coefficients driving the dynamics of σ are bounded on $[0,T]$. For simplicity, we will use the following notations: $V_t = \sigma_t^2 = \sigma_0^2 + \int_0^t f_s ds + \int_0^t g_s dW_s$ and $v_t = \sigma_0^2 + \int_0^t g_s dW_s$. It is easy to see from Lemma 2.1 that V and v both satisfy Assumption 3 with $\gamma^V = \gamma^v = 1$ and $C_\gamma^V = C_\gamma^v$. Now, since

$$h^{-1/2}\mathbb{E}\left|\int_0^T K_h(t-\tau)\int_\tau^t f_s ds dt\right| \leq \sup_{s\in[0,T]} |f_s| h^{-1/2}\int_0^T |K_h(t-\tau)||t-\tau| dt,$$

which is $O_p(h^{1/2}) = o_p(1)$, we can conclude that the drift term of V has a negligible contribution to the final error. Therefore, it suffices to work with the process v and only to consider the weak convergence of $\bar{I}_h := h^{-1/2}\left(\int_0^T K_h(t-\tau)(v_t - v_\tau)dt\right)$. For the sake of clarity, we will first assume a one-sided kernel function (i.e., $K(x) = 0$ for all $x < 0$), so that $\bar{I}_h = h^{-1/2}\left(\int_\tau^T K_h(t-\tau)(v_t - v_\tau)dt\right) =: I_h$. Applying the integration by parts formula, we have that

$$I_h = -h^{-1/2} L\left(\frac{T-\tau}{h}\right)(v_T - v_\tau) + h^{-1/2}\int_\tau^T L\left(\frac{t-\tau}{h}\right) g_t dW_t =: R + S,$$

where $L(t) = \int_t^\infty K(u)du$ so that $\frac{d}{dt}(L((t-\tau)/h)) = -K_h(t-\tau)$. Since our assumptions on K imply that $x^{1/2}L(x) \to 0$, as $x \to \infty$, we have $R = o_P(1)$. For the other term S, let us consider the following approximation $\widetilde{S} := h^{-1/2} g_\tau \int_\tau^T L\left(\frac{t-\tau}{h}\right) dW_t$, and note that $S - \widetilde{S} = o_P(1)$, since by Assumption 4, $\int_0^\infty L^2(x)dx < \infty$ and

$$\begin{aligned}\mathbb{E}\left[\left(S-\widetilde{S}\right)^2\right] &= \frac{1}{h}\left(\int_\tau^{\tau+\sqrt{h}} + \int_{\tau+\sqrt{h}}^T\right) L^2\left(\frac{t-\tau}{h}\right)\mathbb{E}\left[(g_t - g_\tau)^2\right] dt \\ &\leq \sup_{t\in[\tau,\tau+\sqrt{h}]}\mathbb{E}\left[(g_t-g_\tau)^2\right]\|L^2\|_1 + 4\|g^2\|_\infty \int_{1/\sqrt{h}}^\infty L^2(s)ds,\end{aligned}$$

which is clearly $o(1)$, as $h \to 0$. We also observe that conditional on $\mathcal{F}_\tau$, $\tilde{S}$ is Gaussian with mean 0 and variance:

$$g_\tau^2 h^{-1} \int_\tau^T L^2\left(\frac{t-\tau}{h}\right) dt = g_\tau^2 \int_0^{\frac{t-\tau}{h}} L^2(s)ds \to g_\tau^2 \iint K(x)K(y)C_1(x,y)dxdy.$$

Therefore, $\tilde{S}|\mathcal{F}_\tau \to_D \mathcal{N}(0,\delta_2^2)$, where $\delta_2^2 = g_\tau^2 \iint K(x)K(y)C_\gamma(x,y)dxdy$.

We now consider the general two-sided kernel case. To this end, let $\bar{L}(t) = \int_t^\infty K(u)du\mathbf{1}_{\{t>0\}} - \int_{-\infty}^t K(u)du\mathbf{1}_{\{t\le 0\}}$ and note that, by the integration by parts formula, $\bar{I}_h = h^{-1/2}\int_0^T K_h(t-\tau)(v_t - v_\tau)dt$ is such that

$$\begin{aligned}\bar{I}_h &= -h^{-1/2}\bar{L}\left(\frac{T-\tau}{h}\right)(v_T - v_\tau) + h^{-1/2}\int_0^\tau \bar{L}\left(\frac{t-\tau}{h}\right) g_t dW_t + I_h + o_P(1)\\ &=: \bar{R} + \bar{S} + I_h + o_P(1).\end{aligned}$$

Same as in the one-sided kernel case, $\bar{R} = o_P(1)$ and $I_h = \widetilde{S} + o_P(1)$. For $\bar{S}$, we consider the following approximation:

$$\widetilde{\bar{S}} := h^{-1/2} g_\tau \int_0^\tau \bar{L}\left(\frac{t-\tau}{h}\right) dW_t = h^{-1/2} g_\tau \left(\int_0^{\tau-\sqrt{h}} + \int_{\tau-\sqrt{h}}^\tau\right) \bar{L}\left(\frac{t-\tau}{h}\right) dW_t =: \widetilde{\bar{S}}_1 + \widetilde{\bar{S}}_2.$$

We still have $\bar{S} - \widetilde{\bar{S}} = o_P(1)$. It is also true that $\widetilde{\bar{S}}_1 = o_P(1)$, as $h \to 0$, which can be justified by considering its second moment. Therefore, we have

$$\bar{I}_h = \widetilde{\bar{S}}_2 + \widetilde{S} + o_P(1) = h^{-1/2} g_{\tau-\sqrt{h}} \int_{\tau-\sqrt{h}}^T \bar{L}\left(\frac{t-\tau}{h}\right) dW_t + o_P(1) =: \widetilde{\bar{I}}_h + o_P(1),$$

where the second equality holds since $\widetilde{\bar{S}}_2 + \widetilde{S} - \widetilde{\bar{I}}_h = o_P(1)$, which again can be justified by considering the second moment and Cauchy-Schwarz' inequality. To conclude (6.3), note that, by conditioning on $\mathcal{F}_{\tau-\sqrt{h}}$,

$$\mathbb{E}\left[\exp\left(iu\widetilde{\bar{I}}_h\right)\right] = \mathbb{E}\left[\exp\left(-\frac{u^2 g_{\tau-\sqrt{h}}^2}{2}\int_{-h^{-1/2}}^{\frac{T-\tau}{h}} \bar{L}^2(s)\, ds\right)\right],$$

which converges to $\mathbb{E}\left[\exp\left(-\frac{u^2 g_\tau^2}{2}\iint K(x)K(y)C(x,y)dxdy\right)\right]$ and we conclude (6.3).

(2) In the whole proof, the superscript (Z) means that the quantity corresponds to process Z, while quantities without such a superscript corresponds to the process σ^2. Let us start by noting that, since Z is a Gaussian process, $h^{-\gamma/2}\left(\int_0^T K_h(t-\tau)(Z_t - Z_\tau)dt\right) \to_D \bar{\delta}_2^{1/2} N(0,1)$, where $\bar{\delta}_2 = L^{(Z)}(\tau)\iint K(x)K(y)C_\gamma^{(Z)}(x,y)dxdy$. Now, for any $\epsilon \in (0, \min(\tau, T-\tau))$, and for any $t \in (\tau-\epsilon, \tau+\epsilon)$, there exists $s_t \in (\min(t,\tau), \max(t,\tau))$, such that $\sigma_t^2 - \sigma_\tau^2 = f'(Z_\tau)(Z_t - Z_\tau) + \frac{1}{2}f''(Z_{s_t})(Z_t - Z_\tau)^2$. Then, $I := \int_0^T K_h(t-\tau)(\sigma_t^2 - \sigma_\tau^2)dt$ is such that

$$I = \int_{\tau-\epsilon}^{\tau+\epsilon} K_h(t-\tau)[f'(Z_\tau)(Z_t - Z_\tau) + \frac{1}{2}f''(Z_{s_t})(Z_t - Z_\tau)^2]dt + o(h^{\gamma/2}).$$

For the second term, once we select ϵ small enough such that $\mathbb{E}[(f''(Z_t))^2] < M^2$ and $\mathbb{E}[(Z_t - Z_\tau)^4] \le M|t-\tau|^\gamma$ for all $t \in (\tau-\epsilon, \tau+\epsilon)$, we have that

$$\mathbb{E}\left|\int_{\tau-\epsilon}^{\tau+\epsilon} K_h(t-\tau)f''(Z_{s_t})(Z_t - Z_\tau)^2 dt\right| \le 3M^2 \int_{\tau-\epsilon}^{\tau+\epsilon} |K_h(t-\tau)||t-\tau|^\gamma,$$

which is $O(h^\gamma) = o(h^{\gamma/2})$. Now for the first term, we have

$$h^{-\gamma/2}\int_{\tau-\epsilon}^{\tau+\epsilon} K_h(t-\tau)[f'(Z_\tau)(Z_t - Z_\tau)]dt \to_D f'(Z_\tau)\bar{\delta}_2^{1/2} N(0,1).$$

where the standard normal $N(0,1)$ appearing above is independent from Z_τ. Indeed, $(X, Y(h)) := (Z_\tau, h^{-\gamma/2}\int_{\tau-\epsilon}^{\tau+\epsilon} K_h(t-\tau)(Z_t - Z_\tau)dt)$ is bi-variate normal for all $h > 0$ and, thus, whenever the limit $(X, Y(h)) \to (X, Y)$ exists, (X, Y) is bivariate normal variable. There exist $\alpha(h)$ and $\beta(h)$ such that $Y(h) = \alpha(h)X + \beta(h)Z(h)$, such that X is independent with $Z(h)$ and $Z(h) \to_D N(0,1)$, as $h \to 0$. Note that $\alpha(h)$ and $\beta(h)$ are given by

$$\alpha(h) = \frac{\mathbb{E}[XY(h)]}{\mathbb{E}[X^2]}, \quad \beta^2(h) = \mathbb{E}[Y^2(h)] - \alpha^2(h)\mathbb{E}[X^2].$$

By our assumption, we have $\mathbb{E}[XY(h)] = o(1)$ and, thus, $\alpha(h) = o(1)$, while $\beta^2(h) = L^{(Z)}(\tau)\iint K(x)K(y)C_\gamma^{(Z)}(x,y)dxdy + o(1)$. With such representations, we have:

$$f'(X)Y(h) = \alpha(h)f'(X)X + \beta(h)f'(X)Z(h) = o_p(1) + \beta(h)f'(X)Z(h),$$

which converges to $\beta f'(X)Z$. ■

Supplement Material

Technical Proofs and Additional Related Results: We provide the proofs of some key lemmas and propositions that were omitted in this article for sake of space.

References

Alvarez, A., Panloup, F., Pontier, M. & Savy, N. (2012), 'Estimation of the instantaneous volatility', *Statistical inference for stochastic processes* **15**(1), 27–59.

Barndorff-Nielsen, O. & Veraart, A. (2009), 'Stochastic volatility of volatility in continuous time', *CREATES Research Paper* **25**.

Epanechnikov, V. (1969), 'Non-parametric estimation of a multivariate probability density', *Theory of Probability & Its Applications* **14**(1), 153–158.

Fan, J. & Wang, Y. (2008), 'Spot volatility estimation for high-frequency data', *Statistics and its Interface* **1**(2), 279–288.

Figueroa-López, J. & Li, C. (2018), 'Supplement to "optimal kernel estimation of spot volatility of stochastic differential equations"'.

Foster, D. & Nelson, D. (1994), ‘Continuous record asymptotics for rolling sample variance estimators’.

Jacod, J. & Protter, P. (2012), *Discretization of Processes*, Springer-Verlag Berlin Heidelberg.

Jones, M., Marron, J. & Sheather, S. (1996), ‘A brief survey of bandwidth selection for density estimation’, *Journal of the American Statistical Association* **91**(433), 401–407.

Kristensen, D. (2010), ‘Nonparametric filtering of the realized spot volatility: A kernel-based approach’, *Econometric Theory* **26**(01), 60–93.

Mancini, C., Mattiussi, V. & Renò, R. (2004), ‘Spot volatility estimation using delta sequences’, *Scandinavian Actuarial Journal* **2004**(1), 42–52.

Ossiander, M. & Waymire, E. (1989), ‘Certain positive-definite kernels’, *Proceedings of the American Mathematical Society* **107**(2), 487–492.

Park, B. & Marron, J. (1990), ‘Comparison of data-driven bandwidth selectors’, *Journal of the American Statistical Association* **85**(409), 66–72.

Samorodnitsky, G. & Taqqu, M. (1994), *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York.

Vetter, M. (2015), ‘Estimation of integrated volatility of volatility with applications to goodness-of-fit testing’, *Bernoulli* **21**(4), 2393–2418.

Zhang, L., Mykland, P. & Aït-Sahalia, Y. (2005), ‘A tale of two time scales’, *Journal of the American Statistical Association* **100**(472).